European Research Studies
Volume VI, Issue (3-4), 2003, pp. 121-130

A Semiparametric Frequency Domain Approach of Modelling the Real Output with Fractional Integration

Luis A. Gil-Alana[*1]

Abstract

We propose in this article the use of the Quasi Maximum Likelihood Estimate of Robinson (QMLE, 1995a) for estimating the fractional differencing parameter in the real output and in the growth rate series of France, Italy and the U.K. This method is semiparametric and is robust to the different types of I(0) disturbances. The results indicate that the orders of integration of the real output are, in the three series, higher than 1 but smaller than 1.5, implying that their first differences are stationary but with long memory behaviour. However, the growth rates appear to be I(0) for the three countries. This may appear contradictory at first sight but it is explained by the non-linear log-transform involved. As a conclusion, the results show that the log-transformed series should be the data used in a multivariate framework when relating the real output with other variables in a cointegrating system.

Keywords: Fractional integration; Semiparametric estimation; Mean reversion.
JEL Classification: C22.

1. Introduction

Despite the vast literature on modelling macroeconomic time series, there still exists little consensus about the appropriate way of modelling the real output. Deterministic approaches based on linear (or quadratic) functions of time were shown to be inappropriate in many cases, and stochastic approaches based on first (or second) differences of the data were proposed, especially after the seminal paper of Nelson and Plosser (1982), who, following the work and ideas of Box and Jenkins (1970), showed that many series could be specified in terms of unit roots. However, in the last few years, the unit root model has been discouraged in favour of a much more general type of long memory models, called fractionally integrated, in which the number of differences is fractional rather than an integer value. Thus, for example, in a recent paper, Candelon and Gil-Alana (2003) show that fractionally ARIMA (ARFIMA) models can better replicate the business cycle features of the US output than other linear and non-linear models.

[*] The author gratefully acknowledges financial support from the Government of Navarra, throughout the Grant "Ayudas de Formacion y de Investigacion y Desarrollo".

[1] Luis A. Gil-Alana, Universidad de Navarra, Facultad de Ciencias Economicas, Edificio Biblioteca, Entrada Este, E-31080 Pamplona, SPAIN, Phone:00 34 948 425 625, Fax:00 34 948 425 626, Email: alana@unav.es

In this article we show that the real output in France, Italy and the UK may be specified in terms of I(d) statistical models. For this purpose, we make use of the Quasi Maximum Likelihood Estimate of Robinson (QMLE, 1995a). This method is semiparametric and therefore does not require to make parametric modelling assumptions about the underlying I(0) disturbances. The method is briefly described in Section 2. In Section 3, it is applied to the annual series of the real GDP, while Section 4 contains some concluding comments.

2. The Quasi Maximum Likelihood Estimate of Robinson (1995a)

For the purpose of the present paper, we define an I(0) process as a covariance stationary process with spectral density function that is positive and finite at the zero frequency. In this context, we say that x_t is $I(d)$ if:

$$(1-L)^d x_t = u_t, t = 1, 2, ..., \tag{1}$$

$$x_t = 0, t \leq 0, \tag{2}$$

where L is the lag operator ($Lx_t = x_{t-1}$) and u_t is $I(0)$. If $d > 0$ in (1), x_t is said to be a long memory process, so-called because of the strong association between observations widely separated in time. The macroeconomic literature has stressed the cases of integer d, usually 0 or 1 (unit roots). However, as shown by Adenstedt (1974) and Taqqu (1975), d can also be a real number. This type of processes was introduced by Granger and Joyeux (1980), Granger (1980, 1981) and Hosking (1981), and empirical applications based on fractional models like (1) in macroeconomic time series are amongst others the papers of Diebold and Rudebusch (1989), Baillie and Bollerslev (1994), Gil-Alana and Robinson (1997) and Gil-Alana (2000).

To determine the appropriate degree of integration in a given raw time series is important from both economic and statistical viewpoints. If $d = 0$, the series is covariance stationary and possesses 'short memory', with the autocorrelations decaying fairly rapid. If d belongs to the interval (0, 0.5), x_t is still covariance stationary, however, the autocorrelations take much longer time to disappear than in the previous case. If $d \in [0.5, 1)$, the series is no longer covariance stationary, but it is still mean reverting, with the effect of the shocks dying away in the long run. Finally, if $d \geq 1$, x_t is nonstationary and non-mean reverting. Thus, the fractional differencing parameter d plays a crucial role in describing the persistence in the time series behaviour: higher d is, higher will be the degree of association between the observations.

There exist many approaches for estimating and testing the fractional differencing parameter. Some of them are parametric, in which the model is specified up to a finite number of parameters, (e.g., Fox and Taqqu, 1986; Dahlhaus, 1989; Sowell, 1992; Robinson, 1994a; Tanaka, 1999; etc.). However, on estimating with parametric approaches, the correct choice of the model is important. If it is misspecified, the estimates are liable to be inconsistent. In fact, misspecification of the short-run components of the series may invalidate the estimation of the long run parameter. Thus, there may be some advantages on estimating d with semiparametric techniques. We propose here the use of the Quasi Maximum Likelihood Estimate of Robinson (QMLE, 1995a).[2]

It is basically a 'Whittle estimate' in the frequency domain, considering a band of frequencies that degenerates to zero. The estimate is implicitly defined by:

$$\hat{d} = \arg\min_d \left(\log \overline{C(d)} - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j \right), \tag{3}$$

$$\overline{C(d)} = \frac{1}{m} \sum_{j=1}^{m} I(\lambda_j) \lambda_j^{2d}, \quad \lambda_j = \frac{2\pi j}{T}, \quad \frac{m}{T} \to 0,$$

where $I(\lambda_j)$ is the periodogram of the raw time series, x_t, given by:

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} x_t e^{i\lambda_j t} \right|^2,$$

and $d \in (-0.5, 0.5)$.[3] Under finiteness of the fourth moment and other mild conditions, Robinson (1995a) proved that:

$$\sqrt{m}(\hat{d} - d_o) \to_d \quad N(0,1/4) \quad as \quad T \to \infty,$$

where d_0 is the true value of d and with the only additional requirement that $m \to \infty$ slower than T. In fact, m must be smaller than $T/2$ to avoid aliasing effects.[4] A multivariate extension of this estimation procedure can be found in Lobato (1999). There also exist other semiparametric procedures for estimating the fractional differencing parameter, for example, the log-periodogram regression estimate (LPE), initially proposed by Geweke and Porter-Hudak (1983) and modified later by Kónsch (1986) and Robinson (1995b) and the averaged periodogram estimate (APE) of Robinson (1994b). We have decided to use in this article the QMLE, firstly because of its computational simplicity, though the computation of the LPE is also quite simple. Note, however that using the QMLE, we do not need to employ any additional user-chosen numbers in the estimation (as is the case with the LPE and the APE). Also, we do not have to assume Gaussianity in order to obtain an asymptotic normal distribution, the QMLE being more efficient than the LPE.[5]

3. Testing the order of integration in the real output series

The time series data analysed in this section correspond to the real GDP in France, Italy and the United Kingdom, annually, for the time period 1871-2000, obtained from the International Monetary Fund (IMF) database.

---

2 In the terminology of Marinucci and Robinson (1999), the I(d) models correspond to a type II fractional Brownian motion, which is not the model for which the QMLE was initially designed for. However, it remains valid as long as d belongs to the stationary region.

3 Velasco (1999a,b) and more recently Shimotsu (2000) and Shimotsu and Phillips (2000) have showed that the fractionally differencing parameter can also be consistently semiparametrically estimated in nonstationary contexts by means of tapering.

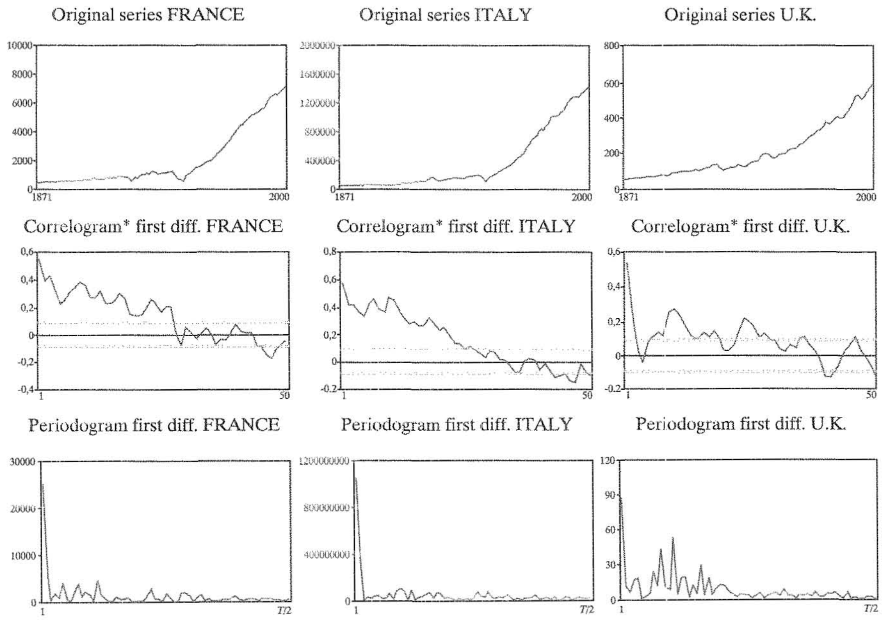
4 The exact requirement is that $(1/m)+((m^{1+2\alpha}(\log m)^2)/(T^{2\alpha}))\to 0$ as $T\to\infty$, where α is determined by the smoothness of the spectral density of the short run component. In case of a stationary and invertible ARMA, α may be set equal to 2 and the condition is then $(1/m)+((m^5(\log m)^2)/(T^4))\to 0$ as $T\to\infty$.

5 Velasco (2000) showed that Gaussianity is not necessary for the LPE either.

We display in the upper part of Figure 1 the plots of the three time series. We observe a clear nonstationary pattern in all of them, with values increasing across the sample.[6] The second and third raws of the figure display respectively the correlograms and the periodograms of the first differenced data. Starting with the correlograms, we observe significant values, some of them relatively far away from 0, which may suggest that first differencing is not sufficient to eliminate the long memory property of the series. This is corroborated by the periodograms, which show a significant value at the smallest frequency in the three series, giving further support to the hypothesis that long memory is still present in the data. Note that though the periodogram is not a consistent estimate of the spectral density function, it is a very useful tool when determining if a time series is short or long memory.[7]

Figure 1

Original series and the correlograms and periodograms of the first differenced data



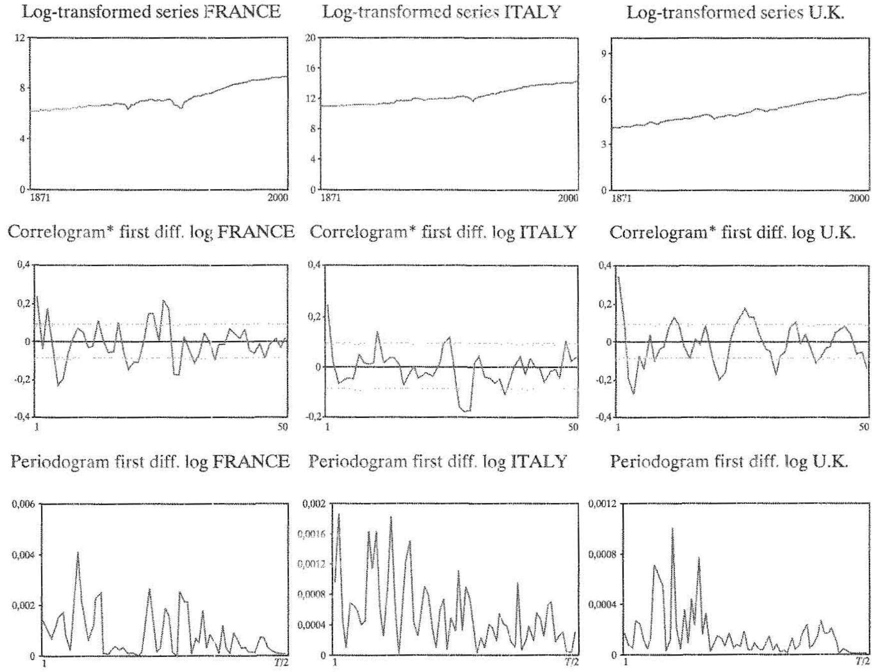
* *The large sample standard error under the null hypothesis of no utocorrelation is $1/\sqrt{T}$ or roughly 0.088.*

---

6 A potential break at World War II may occur, especially for the cases of France and Italy. However, its study is out of the scope of the present work. Moreover, structural change in the context of fractional integration with semiparametric methods has not yet been studied.

7 If a process is I(0) (i.e., short memory), its spectral density function $f(\lambda)$ is positive and finite at the origen, i.e., $0 < f(0) < \infty$. On the other hand, if $d > 0$, (long memory), the spectral density goes to infinity at the origin ($f(0) \rightarrow \infty$) and we should expect the periodogram to replicate this pattern.

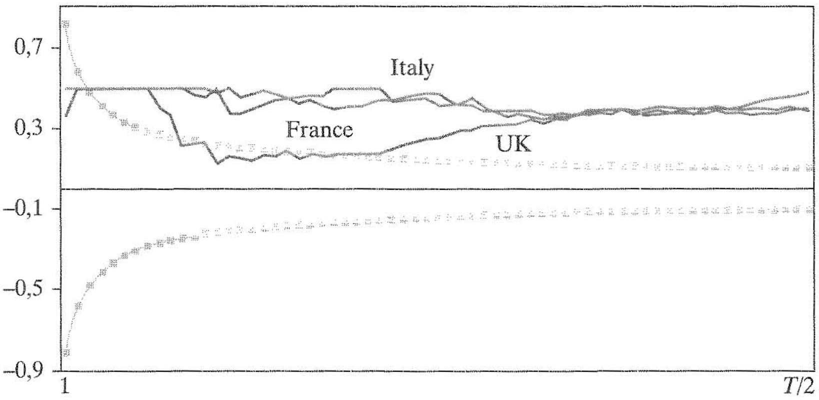
Figure 2
Log-transformed series and the correlograms
and periodograms of the first differences



* *The large sample standard error under the null hypothesis of no autocorrelation is* $1/\sqrt{T}$ *or roughly 0.088.*

Figure 2 is similar to Figure 1 but based on the log-transformed data. Though the series are now smoother than in Figure 1, they are still clearly nonstationary. Similarly to the previous case, the correlograms of the first differences still show significant values at some lags, though the periodograms do not exhibit now a peak at the zero frequency, which may indicate that first differences are now I(0). Because of this different behaviour in the periodograms at the smallest frequency, it is interesting to perform the analysis of the QMLE based on both the original series and the log-transformed data. Since all the time series are nonstationary, the analysis will be carried out based on the first differenced data, adding then 1 to the estimated values of d to obtain the proper orders of integration of the series.

Figure 3
QMLE (Robinson, 1995a) in the first differenced real output series

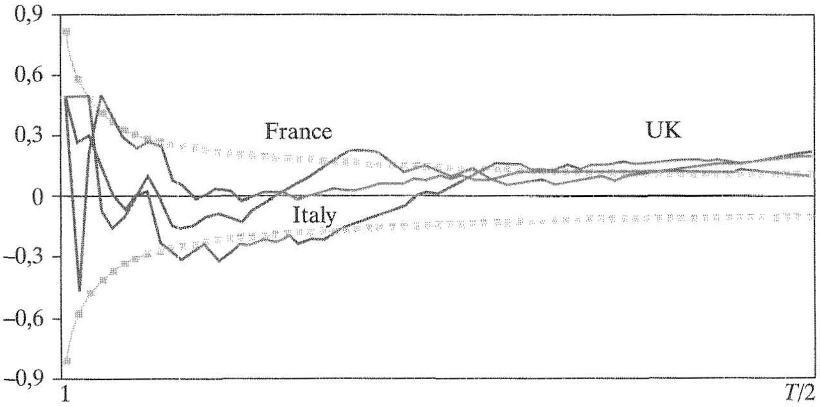


The horizontal axis corresponds to the parameter m while the vertical one refers to the estimate of d.

The results based on the QMLE of Robinson (1995a) are displayed in Figures 3 and 4. We examine the estimates for the whole range of values of m. Some attempts to calculate the optimal bandwidth numbers in long memory contexts have been examined in Delgado and Robinson (1996) and Robinson and Henry (1996). However, in the case of the Whittle estimator (QMLE), the use of optimal values has not been theoretically justified. Other authors, such as Lobato and Savin (1998) use a grid of bandwidth numbers but we have preferred to report the results for the whole range of values of m. Starting with the original data, (Figure 3), we see that the estimates lie between 0.25 and 0.50 for the three countries. The higher estimates are obtained for Italy, followed by France and the UK. In case of Italy, the most stable behaviour is obtained when d is around 0.37. For France, the estimates are slightly below 0.35, and for the UK, d remains stable when it is around 0.29, implying respectively orders of integration of approximately 1.37, 1.35 and 1.29. We have also displayed in the figure the 95% confidence intervals corresponding to the $I(0)$ hypothesis ($I(1)$ on the original series). We see that for all values of m (in case of France and Italy) and for practically all of them (in the UK), the estimates are above the upper bound of the interval, rejecting thus the null hypothesis of $I(0)$ stationarity. In view of these results we can conclude by saying that the three time series are clearly nonstationary and not mean-reverting, while their first differences are covariance stationary but with long memory behaviour. These results suggest that any shock affecting to the levels of the series will be persistent and thus, strong policy actions must be required to bring the variables back to their original levels.

Figure 4

QMLE (Robinson, 1995a) based on the growth rate series



The horizontal axis corresponds to the parameter m while the vertical one refers to the estimate of d.

Next, we look at the behaviour of the growth rates, and perform the same procedure as in Figure 3 but based on the first differences of the logarithmic transformations. The results are displayed in Figure 4. We see here that the estimates are in all cases smaller than in the previous figure, oscillating around 0 if $m \in (0, T/4)$ and slightly higher if $m > T/4$. However, for practically all values of m, the estimates are within the $I(0)$ interval, suggesting that first differences are adequate to eliminate the long memory on the log-transformed data.

The implications of these results are important. Thus, the real output series are $I(d)$ with d higher than 1, implying that not only they are nonstationary but also non-mean reverting, with the effect of the shocks persisting forever. However, the growth rate series are short memory, with the effect of the shocks dying away relatively fast.

These results may appear surprising since we find evidence of long memory for the increments in the real output in the three series but short memory for their corresponding growth rates. However, it should be noted that the analysis is based for each case on a different transformation of the data. Thus, in the first case, we look at $y_t - y_{t-1}$, while in the second case we use $\log y_t - \log y_{t-1}$, this being probably the reason for the differences in the results. On the other hand, this also has some implications for the econometric work. Thus, all the multivariate analysis of the series, which is usually based on cointegration techniques, should be performed on the log-transformed data, which are $I(1)$ variables and do not possess any component of long memory behaviour. On the contrary, if work is based on the original series, the results using cointegration methods may lead to spurious relationships due to the existence of long memory, with d higher than 1 for the individual series.

4. Concluding Comments

We have examined in this article the orders of integration of the annual series of the real output in France, Italy and the UK, along with their corresponding growth rates, by means of using fractionally integrated semiparametric techniques. Using the QMLE procedure of Robinson (1995a) we have shown that the three time series are $I(d)$ with $d > 1$, the orders of integration ranging between 1.25 and 1.40 for the three countries. Thus, the three series are nonstationary, with the increments still possessing a component of long memory behaviour. However, performing the same procedure based on the growth rate series, the results indicate that they are $I(0)$, implying that shocks affecting to the growth rates disappear faster than those affecting to the increments of the levels.

In conclusion, this paper shows that is the log-transformed data the values that should be employed when working in a multivariate model with variables related with the real output, at least for the series examined in the paper. Otherwise, if we decide to work with the data in levels, the series present a high degree of persistence, and any empirical work using cointegration techniques will be invalidated because of the presence of long memory in the differenced series. Other issues, such as the potential presence of structural breaks or the existence of linear trends in the data will be examined in future papers.

References

Adenstedt, R.K., 1974, "On large sample estimation for the mean of a stationary random sequence", *Annals of Statistics* 2, 1095-1107.

Baillie, R.T. and T. Bollerslev, 1994, "Cointegration, fractional cointegration and exchange rate dynamics", *Journal of Finance* 49, 737-745.

Box, G.E.P. and G.M. Jenkins, 1970, *Time series analysis, forecasting and control,* San Francisco, Holden Day.

Candelon, B. and L. A. Gil-Alana, 2003, "Fractional integration and business cycle features", forthcoming in *Empirical Economics.*

Dahlhaus, R., 1989, "Efficient parameter estimation for self-similar process", *Annals of Statistics* 17, 1749-1766.

Delgado, M. and P.M. Robinson, 1996, "Optimal spectral bandwidthfor long memory", *Statistica Seneca* 6, 97-112.

Diebold, F.X. and G.D. Rudebusch, 1989, "Long memory and persistence in aggregate output", *Journal of Monetary Economics* 24, 189-209.

Fox, R. and M.S. Taqqu, 1986, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series", *Annals of Statistics* 14, 517-532.

Geweke, J. and S. Porter-Hudak, 1983, "The estimation and application of long memory time series models", *Journal of Time Series Analysis* 4, 15-39.

Gil-Alana, L.A., 2000, "Mean reversion in the real exchange rates", *Economics Letters* 16, 285-288.

Gil-Alana, L.A. and P.M. Robinson, 1997, "Testing of unit roots and other nonstationary hypotheses in macroeconomic time series", *Journal of Econometrics* 80, 241-268.

Granger, C.W.J., 1980, "Long memory relationships and the aggregation of dynamic models", *Journal of Econometrics* 14, 227-238.

Granger, C.W.J., 1981, "Some properties of time series data and their use in econometric model specification", *Journal of Econometrics* 16, 121-130.

Granger, C.W.J. and R. Joyeux, 1980, "An introduction to long memory time series and fractionally differencing", *Journal of Time Series Analysis* 1, 15-29.

Hosking, J.R.M., 1981, "Modelling persistence in hydrological time series using fractional differencing", *Water Resources Research* 20, 1898-1908.

Kónsch, H., 1986, "Discrimination between monotonic trends and long-range dependence", *Journal of Applied Probability* 23, 1025-1030.

Lobato, I., 1999, "A semiparametric two-step estimator for a multivariate long memory process", *Journal of Econometrics* 90, 129-153.

Lobato, I. and N:E: Savin, 1998, "Real and spurious long memory properties of stock market data", *Journal of Business and Economic Statistics* 16, 261-283.

Marinucci, D. and P.M. Robinson, 1999, "Alternative forms of fractrional Brownian motion", *Journal of Statistical Planning and Inference* 80, 111-122.

Nelson, C.R. and C.I. Plosser, 1982, "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics* 10, 139-162.

Robinson, P.M., 1994a, "Efficient tests of nonstationary hypotheses", *Journal of the American Statistical Association* 89, 1420-1437.

Robinson, P.M., 1994b, "Semiparametric analysis of long memory time series", *Annals of Statistics* 22, 515-539.

Robinson, P.M., 1995a, "Gaussian semiparametric estimation of long range dependence", *Annals of Statistics* 23, 1630-1661.

Robinson, P.M., 1995b, "Log-periodogram regression of time series with long range dependence", *Annals of Statistics* 23, 1048-1072.

Robinson, P.M. and M. Henry, 1996, Bandwidth choice in Gaussian semiparametric estimation of long-range dependence, in P.M. Robinson and M. Rosenblatt eds. *Athens Conferecne on Applied Probability in Time Series Analysis,* Vol.II, New York, 220-232.

Shimotsu, K., 2002, "Exact local Whittle estimation of fractional integration with unknown mean and time trend", Department of Economics, *Discussion Paper* 543, University of Essex.

Shimotsu, K. and P.C.B. Phillips, 2002, "Exact local Whittle estimation of fractional integration", Cowles Foundation, *Discussion Paper* 1367.

Sowell, F., 1992, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models", *Journal of Econometrics* 53, 165-198.

Tanaka, K., 1999, "The nonstationary fractional unit root", *Econometric Theory* 15, 549-582.

Taqqu, M.S., 1975, Weak convergence to fractional Brownian motion and to the Rosenblatt process, Z. Wahrscheinlichkeitstheorie verw. Geb. 31, 287-302.

Velasco, C., 1999a, "Nonstationary log-periodogram regression", *Journal of Econometrics,* 91, 299-323.

Velasco, C., 1999b,. "Gaussian semiparametric estimation of nonstationary time series", *Journal of Time Series Analysis,* 20, 87-127.

Velasco, C., 2000, "Non-Gaussian log periodogram regression", *Econometric Theory* 16, 44-79.